SEST 2020
3rd International Conference on Smart Energy Systems and Technologies (SEST)
7-9 September 2020, Istanbul, Turkey

Paper No: 200

Wasserstein-Distance-Based Temporal Clustering for Capacity-Expansion Planning in Power Systems

Lucas Condeixa
Fabricio Oliveira
Afzal Siddiqui

Aalto University

Γ-opt
Group of Applied Mathematical Modelling And Optimisation

Brief look into the problem and the state-of-art

Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)

Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise

Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise
- Both operational and temporal dimensions can be reduced. We focus in temporal aggregation and its improvement

Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise
- Both operational and temporal dimensions can be reduced. We focus in temporal aggregation and its improvement
- Define capacity via utilization => Load Levels

Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise
- Both operational and temporal dimensions can be reduced. We focus in temporal aggregation and its improvement
- Define capacity via utilization => Load Levels
- Wogrin et al., 2014: Systems states (k-means clustering)

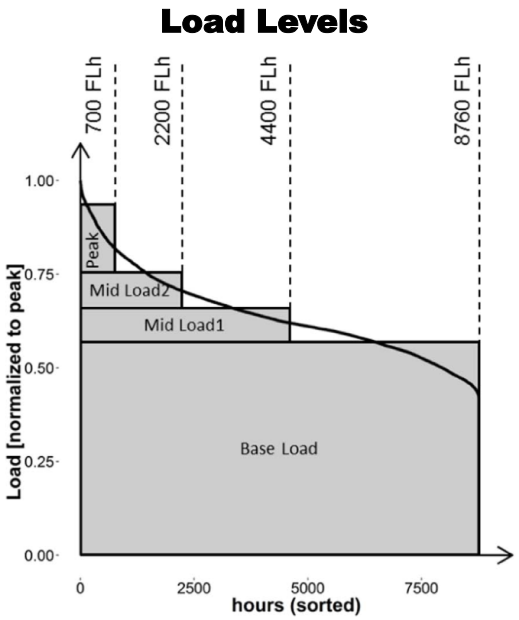
Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise
- Both operational and temporal dimensions can be reduced. We focus in temporal aggregation and its improvement
- Define capacity via utilization => Load Levels
- Wogrin et al., 2014: Systems states (k-means clustering)
- Hierarchical clustering

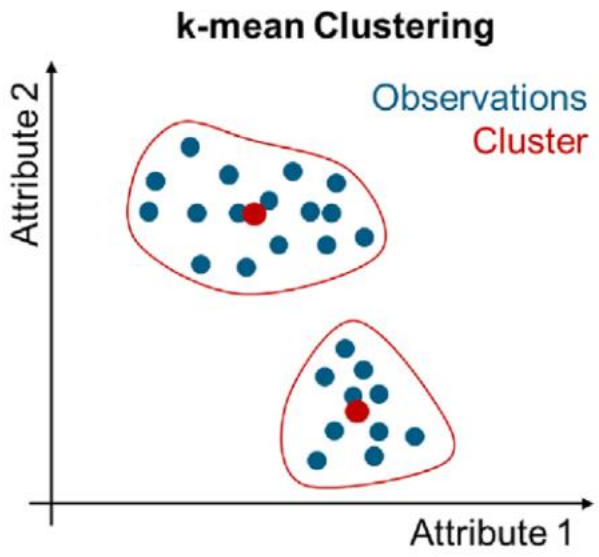
Brief look into the problem and the state-of-art

- Diverse range of decisions to be made when planning the grid => mathematical optimization (Capacity Expansion Planning)
- Increase of VRES penetration suscitates the need of a higher temporal resolution => data selection premise
- Both operational and temporal dimensions can be reduced. We focus in temporal aggregation and its improvement
- Define capacity via utilization => Load Levels
- Wogrin et al., 2014: Systems states (k-means clustering)
- Hierarchical clustering
- Tejada-Arango et al., 2018: Enhanced representative periods

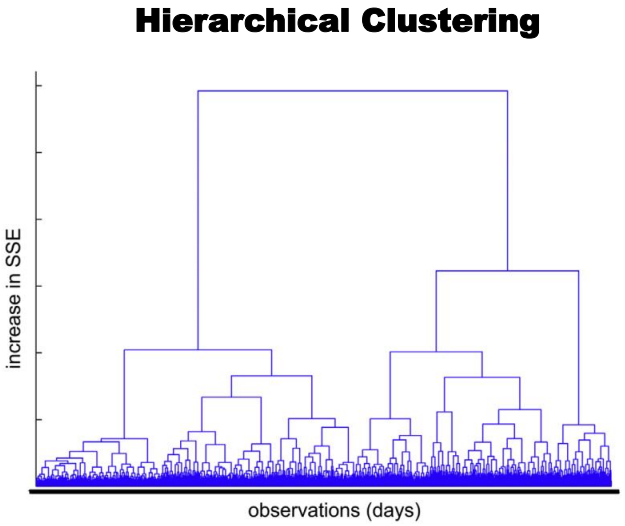
Brief look into the problem and the state-of-art



(+) Quick economical indication
(-) Lack the detailing needed for VRES
(-) Poor at estimating operational outcomes (e.g. electricity prices, start-ups and shut-downs, etc.)



(+) Better overall costs estimation than load levels
(+) Possible to elect the most critical features of a system
(-) Restricted at chronological decisions
(-) Randomness of the path (always dependent on the first centroids sorted)



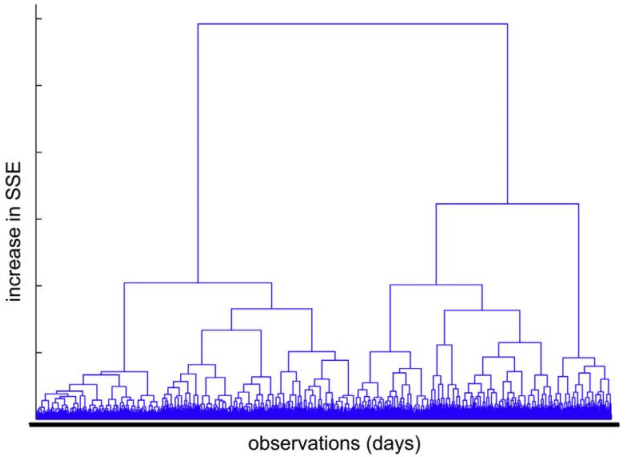
(+) Chronological decisions covered
(+) Consistent path (does not depend on any random classification step)
(-) Weak representation of seasonal patterns

Source: Wogrin et al., 2014; Namacher et al., 2016; Collins et al., 2017; Liu et al., 2018; and Kotzur et al., 2018

Brief look into the problem and the state-of-art

- Chronological decisions will be important to the case studied (i.e. for **storage decisions**)
- We analyze how the choice on the cluster discrepancy measure can improve the Hierarchical Clustering weakness on estimating seasonal patterns

Hierarchical Clustering



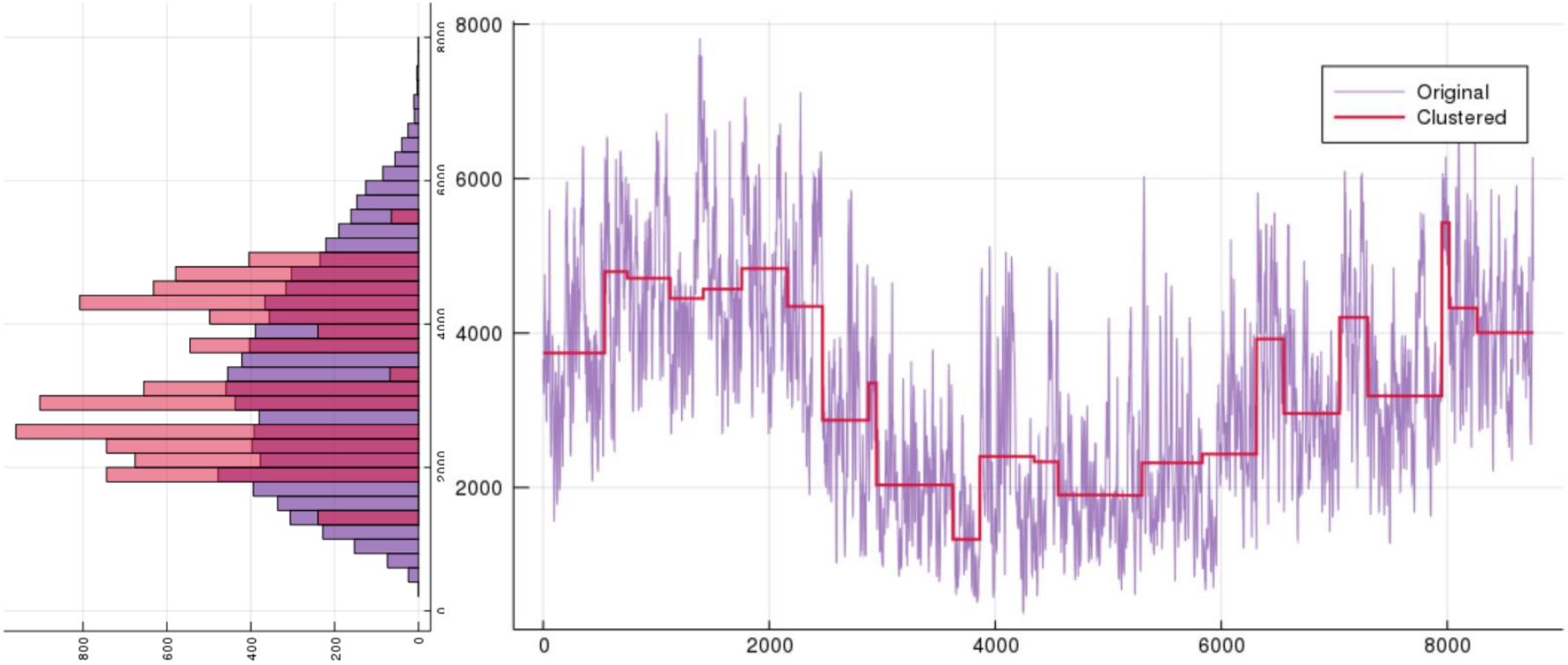
(+) Chronological decisions covered
(+) Consistent path (does not depend on any random classification step)
(-) Weak representation of seasonal patterns

Source: Wogrin et al., 2014; Namacher et al., 2016; Collins et al., 2017; Liu et al., 2018; and Kotzur et al., 2018

Our proposal and why

- The discrepancy metric used to classify different clusters can play a role in the final decision
- Typically, the Euclidean distance (ED) is used to measure distances between clusters
- If we think about the temporal reduction to be performed in the clustering, the idea is similar to what has been done in scenario reduction techniques used in stochastic optimization => **Wasserstein Distance**

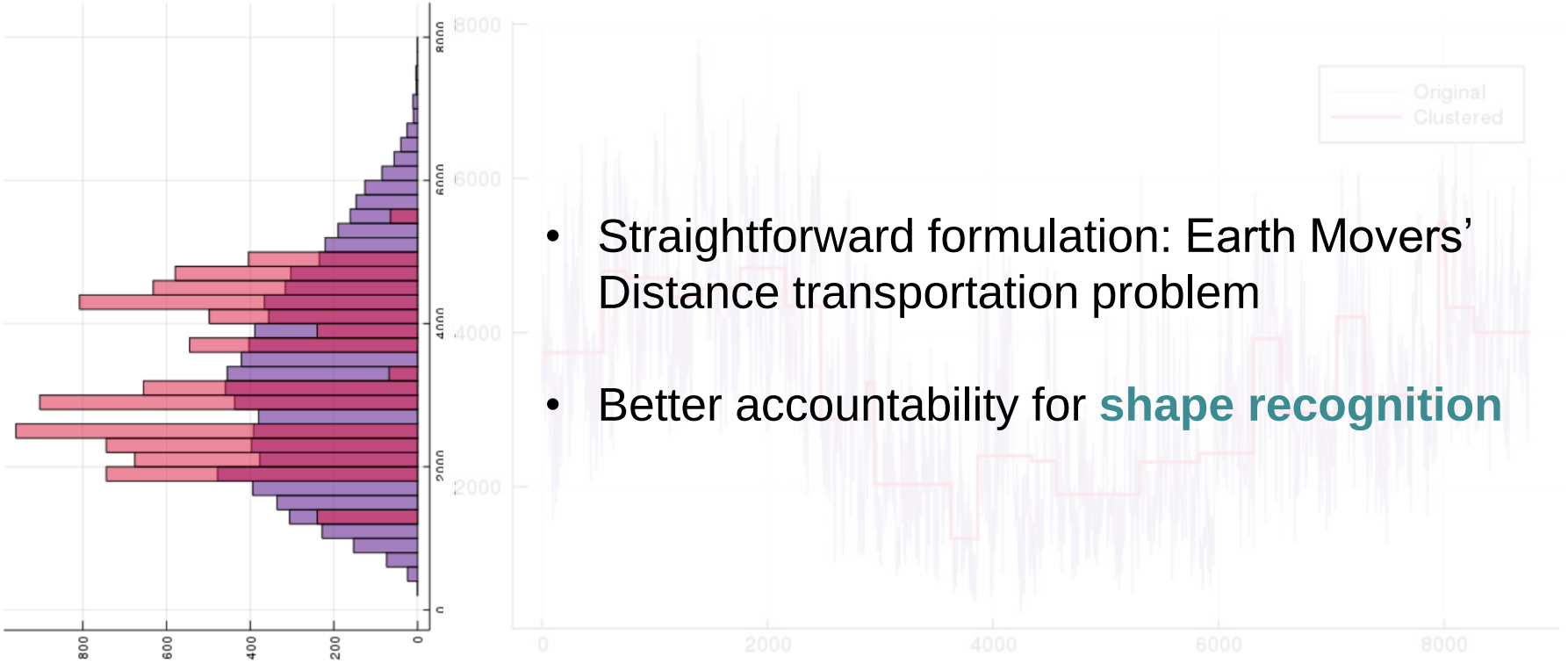
Our proposal and why



$$d^{WD}(Or, Clus) = inf_q \sum_{i,j} q_{ij} \rho(\theta_i, \theta_j)$$

θ_i: Original distribution function
θ_j: Clustered distribution function
ρ : $|i - j|$
q: Quantity transported between i and j

Our proposal and why



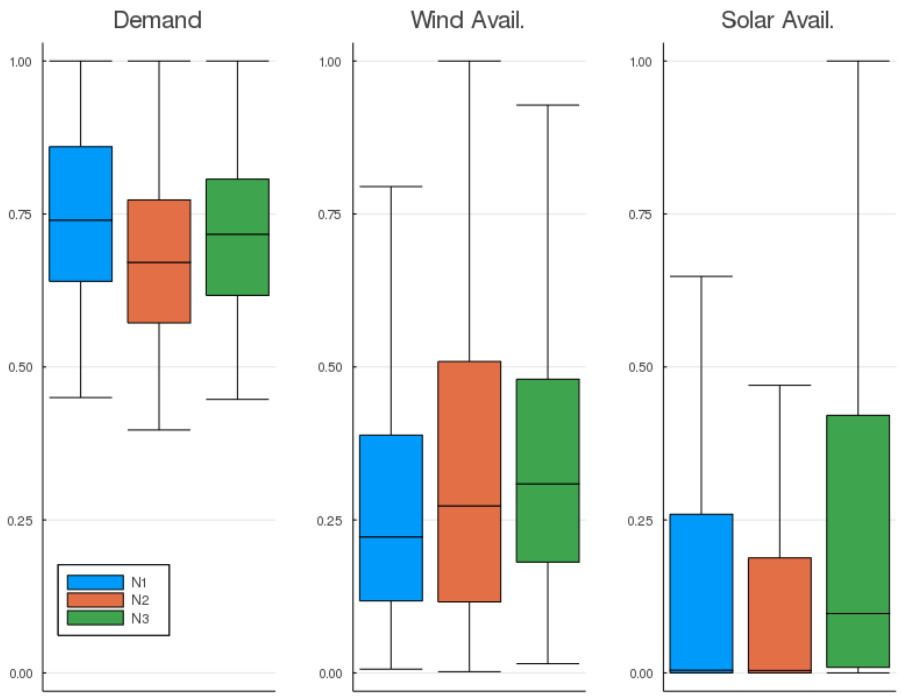
- Straightforward formulation: Earth Movers' Distance transportation problem
- Better accountability for **shape recognition**

$$d^{WD}(Or, Clus) = inf_q \sum_{i,j} q_{ij}\rho(\theta_i, \theta_j)$$

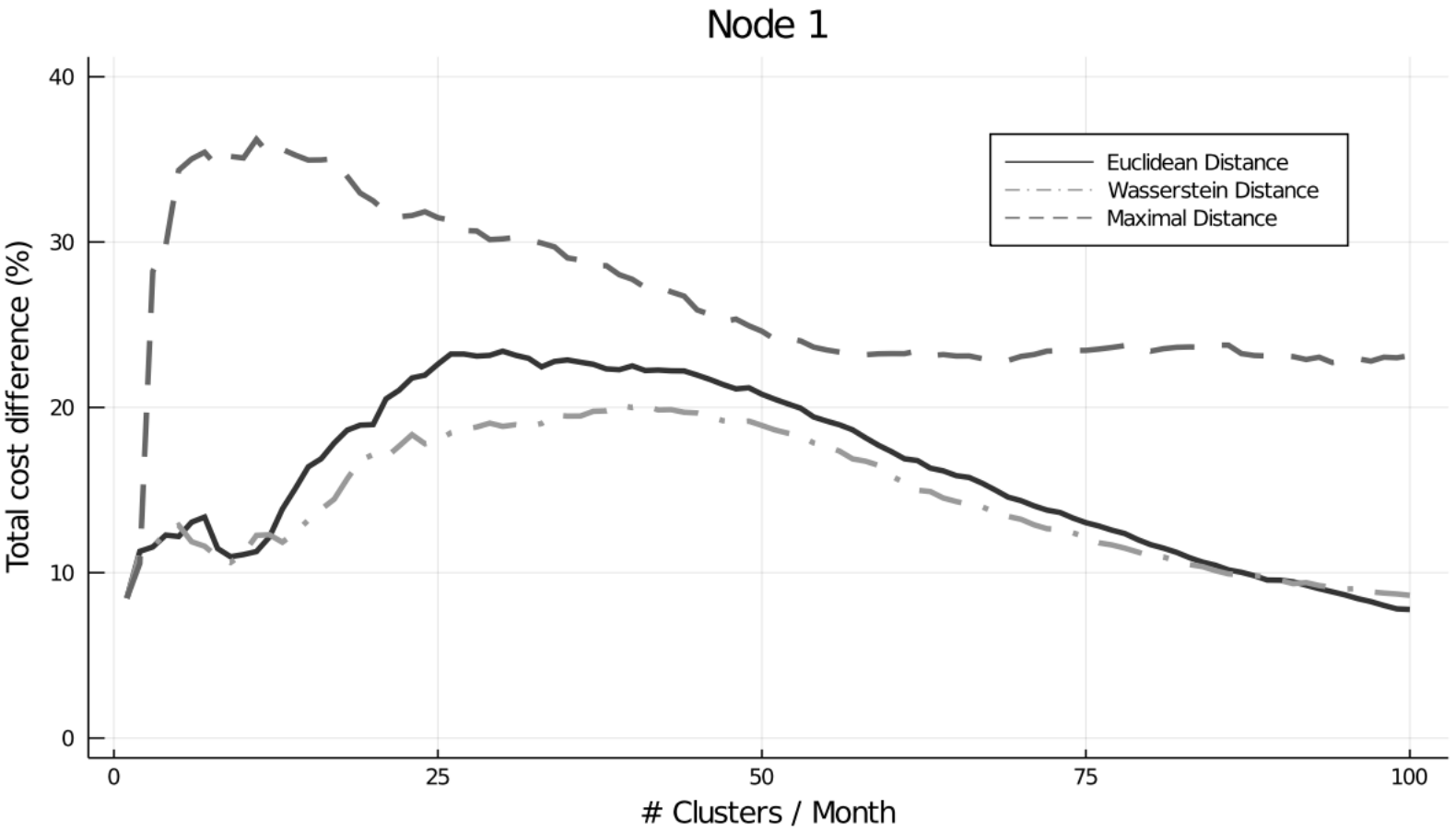
θ_i: Original distribution function
θ_j: Clustered distribution function
ρ : $|i - j|$
q: Quantity transported between i and j

Case study

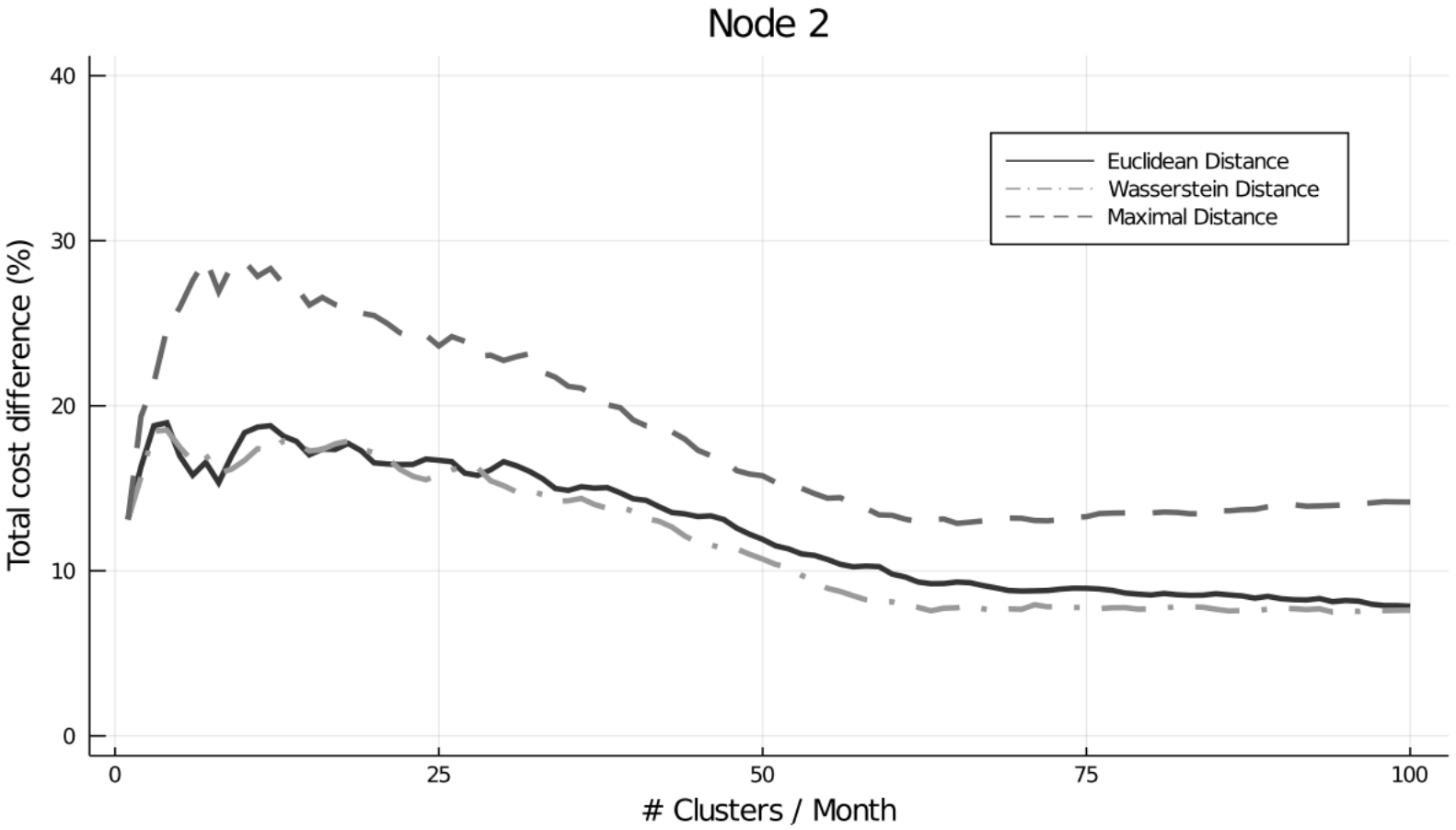
- Use of three different discrepancy measures (Euclidean Distance, Wasserstein Distance, and Maximal Distance) and comparison of the CEP optimization results
- 3 nodes with different conditions: 1 **high demand** (N1), **windy** (N2), and 1 **sunny** (N3)
- CEP model with battery storage
- 50% minimal renewable generation share
- Perfect competition, i.e., producers have no market power to manipulate prices



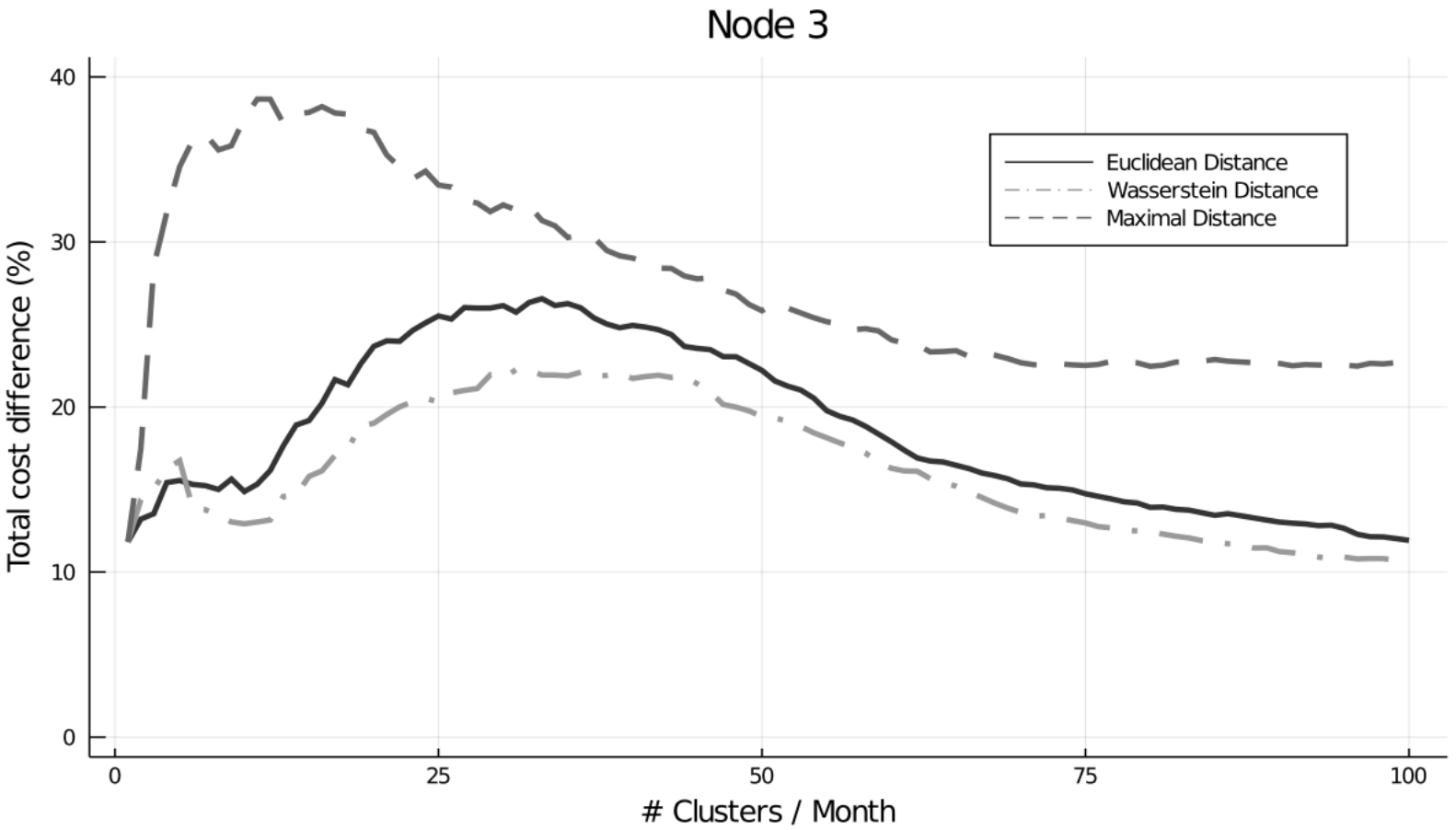
Results and Discussion: Cost analysis



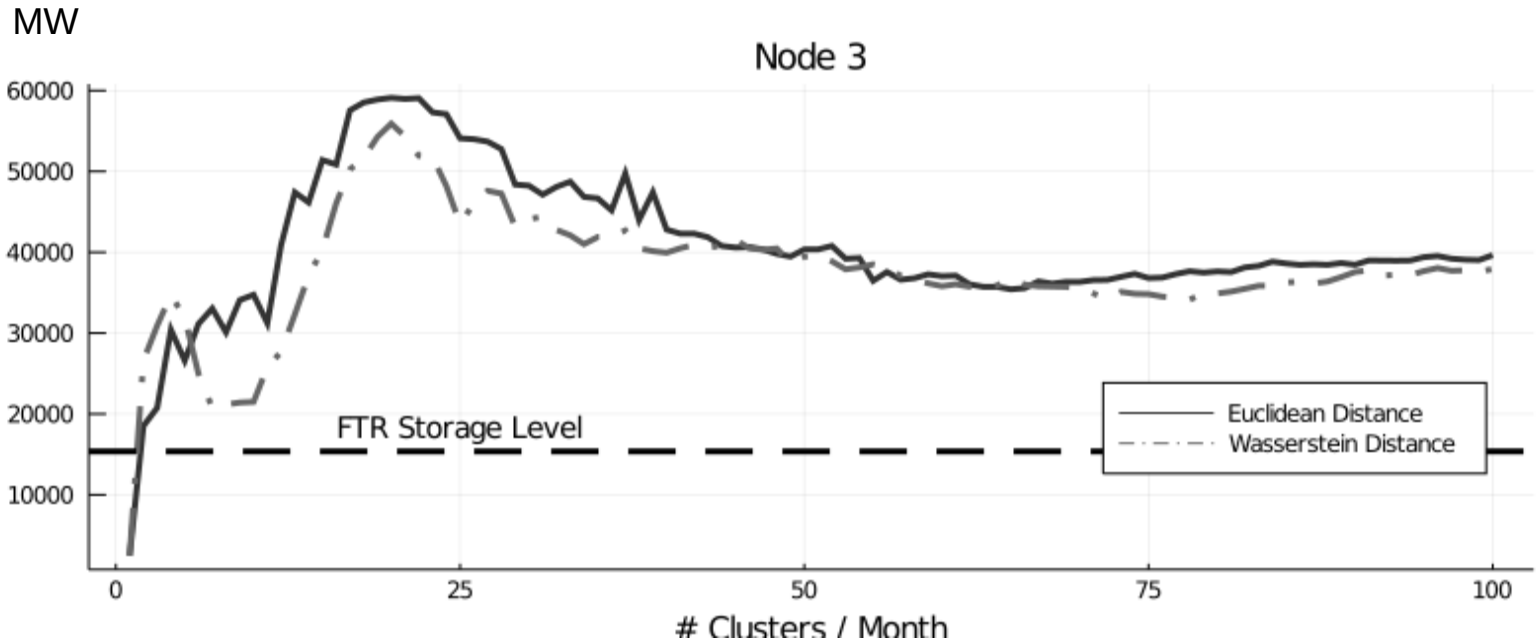
Results and Discussion: Cost analysis



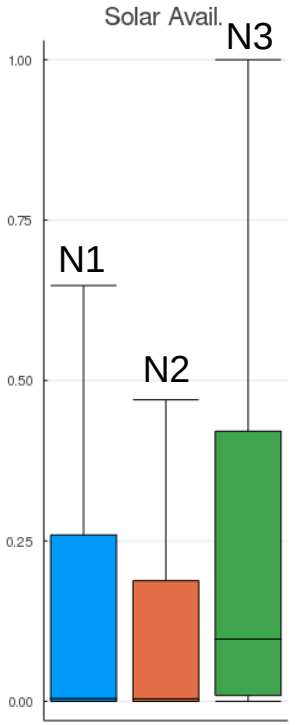
Results and Discussion: Cost analysis



Results and Discussion: Storage capacity (N3)



** Full-Time Resolution*



Solar penetration ↑

Storage Capacity needed ↑

Conclusions

- More **accurate costs** were consistently found for the 3 nodes when using the WD for clustering discrepancy accounting. The estimation of costs is directly related to the cost of capital of an investment. Therefore, by using WD to define clusters, regulators and market players can have a **better estimate on the project's feasibility**

Conclusions

- More **accurate costs** were consistently found for the 3 nodes when using the WD for clustering discrepancy accounting. The estimation of costs is directly related to the cost of capital of an investment. Therefore, by using WD to define clusters, regulators and market players can have a **better estimate on the project's feasibility**
- **Storage capacity estimation could be improved** up to 100% in some cases – nonetheless, still further improvements due

Conclusions

- More **accurate costs** were consistently found for the 3 nodes when using the WD for clustering discrepancy accounting. The estimation of costs is directly related to the cost of capital of an investment. Therefore, by using WD to define clusters, regulators and market players can have a **better estimate on the project's feasibility**
- **Storage capacity estimation could be improved** up to 100% in some cases – nonetheless, still further improvements due
- The operational expenditures (OpEx) were out of this study's scope, nevertheless we have indications that the use of the WD could turn the clustering process more sensible to operational features of the time series analyzed (e.g. solar incidence seasonality)

Conclusions

- More **accurate costs** were consistently found for the 3 nodes when using the WD for clustering discrepancy accounting. The estimation of costs is directly related to the cost of capital of an investment. Therefore, by using WD to define clusters, regulators and market players can have a **better estimate on the project's feasibility**
- **Storage capacity estimation could be improved** up to 100% in some cases – nonetheless, still further improvements due
- The operational expenditures (OpEx) were out of this study's scope, nevertheless we have indications that the use of the WD could turn the clustering process more sensible to operational features of the time series analyzed (e.g. solar incidence seasonality)

Future studies

- Despite the good overall costs estimation while using the WD for clustering, the capacity mix did not benefit as much (also it was not the case for the ED or the MD). We recommend a future work focusing on which type of decisions can benefit better of the WD-based clustering and why

Future studies

- Despite the good overall costs estimation while using the WD for clustering, the capacity mix did not benefit as much (also it was not the case for the ED or the MD). We recommend a future work focusing on which type of decisions can benefit better of the WD-based clustering and why
- Different clustering methods can be used to improve decisions such as the storage capacity (as the hierarchical clustering fails to account for long-term storage effects) or the capacity mix

Future studies

- Despite the good overall costs estimation while using the WD for clustering, the capacity mix did not benefit as much (also it was not the case for the ED or the MD). We recommend a future work focusing on which type of decisions can benefit better of the WD-based clustering and why
- Different clustering methods can be used to improve decisions such as the storage capacity (as the hierarchical clustering fails to account for long-term storage effects) or the capacity mix
- An extension of the numerical examples to an integrated multi-nodal system can inform how WD relates to the temporal correlation between different regions with complementary weather conditions, such as the effects of having a region with a high solar incidence connected to another region with high wind availability

Future studies

- Despite the good overall costs estimation while using the WD for clustering, the capacity mix did not benefit as much (also it was not the case for the ED or the MD). We recommend a future work focusing on which type of decisions can benefit better of the WD-based clustering and why
- Different clustering methods can be used to improve decisions such as the storage capacity (as the hierarchical clustering fails to account for long-term storage effects) or the capacity mix
- An extension of the numerical examples to an integrated multi-nodal system can inform how WD relates to the temporal correlation between different regions with complementary weather conditions, such as the effects of having a region with a high solar incidence connected to another region with high wind availability